Analysis of the *Escherichia coli proBA* locus by DNA and protein sequencing

A.H.Deutch[1*], K.E.Rushlow[2] and C.J.Smith[3]

Bethesda Research Laboratories, Inc., P.O. Box 6009, Gaithersburg, MD 20877, USA

Received 6 March 1984; Revised and Accepted 10 July 1984

ABSTRACT

A 2.9 kb DNA fragment carrying the Escherichia coli proBA region, which encodes the first two enzymes of the proline biosynthetic pathway, was subcloned onto an expression plasmid carrying both the bacteriophage lambda P_L promoter (λP_L) and the lambda gene encoding a thermolabile cI repressor protein (cI857). Derepression of the λP_L promoter by thermal inactivation of the cI857 repressor protein resulted in the simultaneous overproduction of the proB (γ-glutamyl kinase) and proA (γ-glutamyl phosphate reductase) gene products. Nucleotide sequence analysis of the proBA locus allowed gene assignments consistent with the NH_2 and COOH-terminal analyses and amino acid compositions of homogeneous preparations of the proB and proA proteins. The contiguous nature of the proB and proA genes suggests that the two genes constitute an operon in which proB precedes proA.

INTRODUCTION

In Escherichia coli, conversion of glutamate to proline is mediated by the sequential action of three enzymes; γ-glutamyl kinase (ATP: L-glutamate 5-phosphotransferase; EC 2.7.2.11), γ-glutamyl phosphate reductase [L-glutamate 5-semialdehyde: $NADP^+$ oxidoreductase (phosphorylating); EC 1.2.1.41], and Δ^1-pyrroline-5-carboxylate reductase (L-proline: $NAD(P)^+$ 5-oxidoreductase; EC 1.5.1.2). These enzymes are the gene products of the proB, proA and proC loci, respectively (1,2). The first enzyme in the proline biosynthetic pathway, γ-glutamyl kinase (GK), catalyzes the ATP-dependent γ-phosphorylation of glutamate to form γ-glutamyl phosphate and ADP (3). The second enzyme of the proline pathway, γ-glutamyl phosphate reductase (GPR), catalyzes the NADPH-dependent reduction of γ-glutamyl phosphate to glutamate 5-semialdehyde and $NADP^+$ (4). L-glutamate 5-semialdehyde undergoes a spontaneous cyclization to form Δ^1-pyrroline-5-carboxylate which is then converted to proline through the action of the third enzyme in the proline pathway, Δ^1-pyrroline-5-carboxylate reductase (2,5).

The primary goal of our research efforts has been to genetically engineer the Gram-negative bacterium, E. coli, to overproduce free amino acids such as

proline. The overproduction of amino acids via fermentation is partially dependent upon overcoming the normal cellular regulatory circuits such as feedback inhibition and repression. Only GK, the first enzyme in the proline biosynthetic pathway, is subject to allosteric feedback inhibition by proline (3). Although it is unclear from the literature whether proB is subject to repression (1,6,10), neither proA nor proC appear to be regulated in this fashion. In a previous publication (7), we reported the nucleotide sequence of the proC gene, and the overproduction, purification to homogeneity and characterization of the corresponding gene product, Δ^1-pyrroline-5-carboxylate reductase. Evidence presented suggested that transcriptional control of the proC gene might occur by either autogenous regulation or by positive activation.

As part of our continuing studies concerning the regulation of the proline biosynthetic pathway, we now report the overproduction, purification to homogeneity and partial characterization of the gene products of the proB and proA loci, as well as the nucleotide sequence of the proBA region. The proB gene we describe encodes a GK enzyme which is resistant to feedback inhibition by proline.

MATERIALS AND METHODS

Enzymes

Restriction endonucleases, T_4-polynucleotide kinase, bacterial-alkaline phosphatase, T_4-DNA ligase, terminal transferase, and E. coli DNA polymerase I (large fragment) were all supplied by Bethesda Research Laboratories, Inc., and were used as suggested by the manufacturer.

Bacterial Strains and Plasmids

E. coli K12 strain K802 (hsdR supE lacY galK metA) and the ColEI hybrid plasmid pLC7-19 were obtained from the E. coli Genetic Stock Center (Yale University). The expression plasmid pGW7 has already been described (7).

Construction and Screening of pGW7-proBA Recombinants

Plasmid pLC7-19 was obtained from the Clarke and Carbon collection of hybrid plasmids (8). We selected for a spontaneous mutation, mapping within the proB locus of pLC7-19, which conferred cellular resistance to 100 µg/ml of DHP. GK activity, from strains harboring plasmids which conferred DHP resistance, was considerably more resistant to the inhibitory effects of proline than the GK activity from wild-type strains (9).

A 3.0 kb PstI DNA fragment, isolated from the modified pLC7-19 plasmid, was identified as encoding both the affected proB locus and the wild-type proA locus (10). The 3.0 kb PstI DNA fragment was subcloned onto the plasmid

pBR322, and the desired recombinant transformants were identified by the nutritional complementation of proline-auxotrophic E. coli. The proline prototrophic transformants were then transferred to lawns of proline-auxotrophic E. coli and tested for the ability to "cross-feed". Transformants which supported the nutritional requirement of the lawn, and which were therefore capable of overproducing proline, were plated on DHP-containing plates and shown to be resistant to the anti-metabolite. Restriction analysis of the isolated plasmid DNA confirmed the insertion of the 3.0 kb PstI DNA fragment. This recombinant plasmid (pBR322-proBA) was the primary source of proBA DNA for sequence analysis and other in vitro manipulations.

The plasmid pBR322-proBA was digested sequentially with PstI and BglII, and the resulting fragments were fractionated by electrophoresis on a 1.2% agarose gel. The 2.9 kb DNA fragment, containing the functionally intact proBA genes, was recovered by electrophoretic elution and purified by precipitation with spermine (11).

The staggered ends of BamHI cleaved pGW7 and purified 2.9 kb proBA DNAs were made flush by the action of E. coli DNA polymerase I (large fragment). The DNAs were ligated together and used to transform strain K802. DHP resistant transformants were selected at 30°C on minimal agar plates containing DHP at 100 ug/ml. All transformants were ampicillin resistant and had the ability to excrete proline and thus "cross-feed" (some in a temperature dependent fashion) proline-requiring strains.

Total cellular protein prepared from the ampicillin and DHP resistant transformants was analyzed by SDS-polyacrylamide gel electrophoresis as described (7). The specific activities of GK and GPR as a function of temperature induction of the λP_L promoter were also measured. Finally, plasmid DNAs containing the 2.9 kb proBA DNA fragment were analyzed for insert-fragment copy number and orientation.

Cell-free Extract Preparation, Enzyme Assays and Enzyme Purification

Cells were suspended in 50mM Tris base-1mM dithiothreitol, pH 7.2, and disrupted by ultrasonic treatment as previously described (9).

The assay for GK activity in crude extracts was performed as described (1) except that the reaction mixture contained the following in a final volume of 0.25ml at pH 7.0: L-glutamate, 50mM; ATP, 10mM; $MgCl_2$, 20mM; L-methionine D,L-sulfoximine, 2.5mM; hydroxylamine HCl, 100 mM; Tris base, 50mM; and enzyme plus water. Inclusion of the glutamine synthetase inhibitor, methionine sulfoximine, greatly decreased background activity and did not interfere with GK activity (9). The assay for GK activity in purified preparations was

dependent upon the addition of GPR (9, see Discussion), and was performed as described (9). GPR activity was measured as described (1) except that the concentration of the substrate, D,L-Δ^1-pyrroline-5-carboxylate, was increased from 1mM to 2.5mM. The synthesis of D,L,-Δ^1-pyrroline-5-carboxylate was accomplished as described (12).

Purification of GK and GPR

GK was purified to homogeneity by sequential chromatography on DEAE-cellulose, Procion Red-agarose, Cibacron Blue-agarose, and Hydroxylapatite (9). Homogenous GPR was prepared by DEAE-cellulose and Hydroxylapatite chromatography as described (9). Both purified enzymes were essentially free of contaminating protein as judged by SDS-polyacrylamide gel electrophoresis.

DNA Sequencing of the 3.0 kb proBA Region

The DNA sequence of the 3.0 kb PstI restriction fragment encoding the proBA genes was determined utilizing both the chemical modification procedure of Maxam and Gilbert (13), and the dideoxy chain-termination method of Sanger et al. (14). The five base-specific chemical modification reactions (G, G + A, A > C, C + T, and C) were performed essentially as described (13). For dideoxy sequencing, DNA restriction fragments generated by digestion of the 3.0 kb PstI fragment with AluI and TaqI were cloned into the SmaI and AccI sites of M13-mp8 (15), respectively. In addition, the entire 3.0 kb PstI fragment was cloned into the PstI site of M13-mp8.

Amino Acid Analysis

Amino acid compositions of the purified GK and GPR proteins were determined as previously described (7). NH_2-terminal amino acid sequencing (GK and GPR) was performed on an updated Beckman 890C sequencer using polybrene (16). COOH-terminal (GPR) analysis was performed as described previously (17) utilizing carboxypeptidase Y.

RESULTS

Overproduction of GK and GPR

The proBA locus was cloned onto the expression plasmid pGW7, to facilitate the production and purification of the proBA gene products. Transformants were identified that overproduced both GK and GPR in response to temperature elevation. One transformant in particular exhibited enzyme activities that were significantly higher than other transformants when the temperature increased to 42°C. Plasmid pGW7-pro$(BA)_2$, isolated from this transformant, was analyzed by restriction enzyme cleavage and was found to contain a tandem repeat of the 2.9 kb proBA DNA fragment inserted at the BamHI site in pGW7 (Figure 1). Transformants which possessed temperature modulated

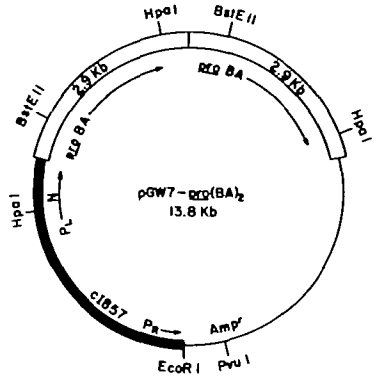

Figure 1. Diagram of the Plasmid pGW7-pro(BA)$_2$ Used to Overproduce the proB and proA Gene Products. The insertion of the direct repeat of the 2.9 kb DNA fragment, encoding the proB (GK) and proA (GPR) gene products, into the BamHI site of the expression plasmid pGW7 is described under Materials and Methods. Thick line, DNA derived from phage λ; thin line, DNA derived from pBR322. Arrows indicate the direction of transcription initiating from the λP_L and λP_R promoters as well as the direction of transcription of the proBA genes.

levels of enzyme activities lower than that obtained in transformants harboring pGW7-pro(BA)$_2$ were found to contain plasmid DNA with a single copy of the 2.9 kb proBA fragment oriented as shown for pGW7-pro(BA)$_2$ in Figure 1. Another class of transformants was found to contain a single copy of the 2.9 kb DNA fragment inserted in pGW7 in the opposite orientation as that found in pGW7-pro(BA)$_2$. Transformants harboring this plasmid construct, pGW7-proBA', were found to possess elevated levels of GK and GPR that were not enhanced following derepression of the λP_L promoter.

The effect of temperature on the specific activities of GK and GPR in cells either containing no plasmid, plasmid pGW7-proBA' or plasmid pGW7-pro(BA)$_2$ is shown in Table 1. A 19-fold and 34-fold increase (compared to wild-type levels, 37°C) was observed in the specific activities of GK and GPR, respectively, in cells harboring plasmid pGW7-proBA' and grown under conditions where the λP_L promoter was repressed (30°C). These increases in specific activities are most likely due to a plasmid copy number effect, with transcription being initiated at the proB promoter (as opposed to transcription originating from an incompletely repressed λP_L promoter), since plasmid pGW7-proBA' does not contain the proBA insert in the proper orientation for transcription from the λP_L promoter. The differential increase observed in the specific activities of the two enzymes can be explained by assuming that translation of the proA mRNA can occur through the

TABLE 1. Specific activities of γ-glutamyl kinase and γ-glutamyl phosphate reductase as a function of increasing temperature[a]

Plasmid[b]	Temperature	Specific Activity[c] γ-glutamyl kinase	Specific Activity[c] γ-glutamyl phosphate reductase
None	30°C	0.592 ± 0.16	2.09 ± 0.50
	37°C	0.498 ± 0.16	1.77 ± 0.65
	42°C	0.525 ± 0.18	1.35 ± 0.60
pGW7-*proBA'*	30°C	9.59 ± 2.2	59.73 ± 10
	37°C	7.30 ± 1.8	47.34 ± 12
	42°C	5.15 ± 1.1	25.96 ± 5
pGW7-*pro*$(BA)_2$	30°C	22.54 ± 2.2	125.0 ± 14
	37°C	25.45 ± 2.5	175.7 ± 20
	42°C	50.41 ± 12.4	613.0 ± 20

[a]Strain K802 with and without plasmid was grown at 30°C in complex broth with and without ampicillin, respectively. At an optical density of 0.65-0.80, at 660 nm, 230ml portions of each culture were incubated at 30°C, 37°C, and 42°C for a period of two hours. Enzyme activity was assayed as described in Materials and Methods.

[b]pGW7-*proBA'* and pGW7-*pro*$(BA)_2$ contain the *proBA* region cloned in opposite orientations. pGW7-*proBA'* contains a single copy of the 2.9 kb *proBA* region in the wrong orientation relative to P_L. pGW7-*pro*$(BA)_2$ contains two copies of the 2.9 kb fragment in a direct tandem repeat oriented as to be placed under the control of the λP_L promoter (see Figure 1).

[c]All values represent the average of three independent experiments ± the standard deviation. γ-Glutamyl kinase activity is expressed as nanomoles product formed per minute per milligram of protein. γ-Glutamyl phosphate reductase activity is expressed as nanomoles NADP reduced per minute per milligram of protein.

binding of ribosomes to either the *proB* or the *proA* ribosomal binding site (see Discussion).

As the temperature of the pGW7-*proBA'* cell culture was increased from 30°C to 42°C both GK and GPR exhibited a decrease in their specific activities of 46% and 57%, respectively. The observed decreases in specific activity is consistent with the *proBA* DNA insert being situated in the opposite orientation as required for transcription from λP_L. The transcriptional activity of RNA polymerase molecules initiating at the *proB* promoter would be opposed by RNA polymerase molecules initiating at the λP_L promoter and transcribing in the opposite direction.

Enzymatic analysis of cells harboring the plasmid pGW7-*pro*$(BA)_2$ and grown at 30°C revealed a 2.2-fold increase in the specific activities of GK and GPR

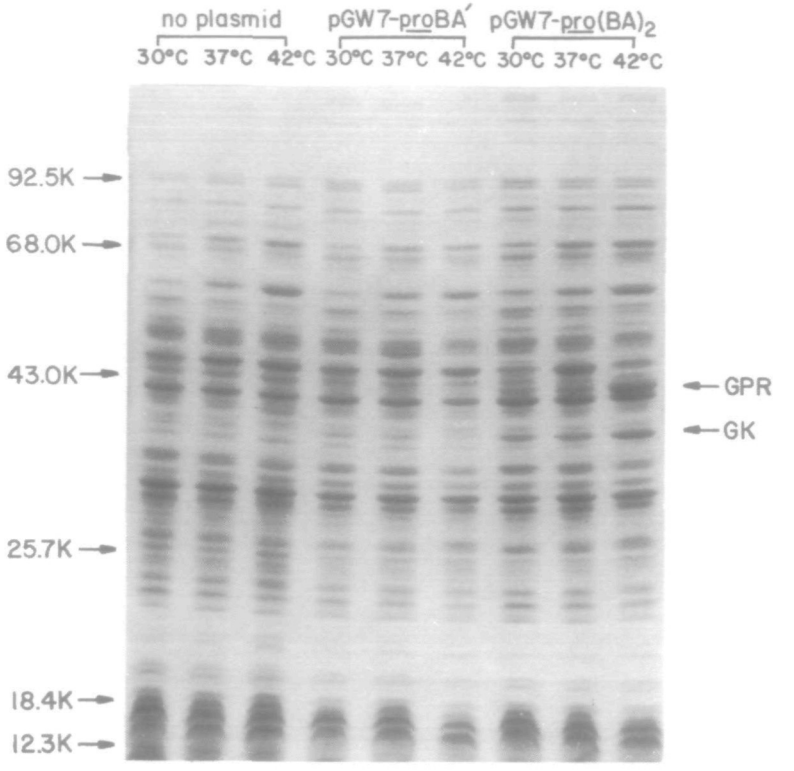

Figure 2. SDS-polyacrylamide Gel Electrophoretic Analysis of Proteins Accumulated During Induction of the λP_L Promoter. E. coli strain K802 with and without plasmid was grown as described in Table 1. Following the two hour incubation at either 30°C, 37°C or 42°C, 0.1 ml aliquots of each of the nine cultures were subjected to centrifugation in an Eppendorf microfuge. The cell pellets were resuspended in electrophoresis sample buffer, boiled for five minutes and electrophoresed on a 7.5%-20% polyacrylamide gradient slab gel containing 0.1% SDS (34). After electrophoresis the protein bands were visualized by staining with Coomassie Brilliant Blue R-250. Arrows on left indicate the position of protein molecular weight markers. Arrows on right indicate the location of γ-glutamyl kinase (GK) and γ-glutamyl phosphate reductase (GPR).

as compared to cells containing pGW7-proBA'. This increase is consistent with the gene dosage effect expected for a plasmid containing a tandem repeat of the proBA DNA fragment. Shifting the temperature of the culture from 30°C to 42°C resulted in an additional 2-fold and 5-fold increase, respectively, in the specific activities of GK and GPR. Overall this corresponded to a respective 100-fold and 350-fold increase in comparison to wild-type cells, and is indicative of control by the λP_L promoter.

Total soluble protein obtained from cells grown at either 30°C, 37°C, or 42°C was fractionated by SDS-polyacrylamide gel electrophoresis and the abundant proteins were detected by staining with Coomassie Brilliant Blue (Figure 2). Cells containing no plasmid and cells containing the plasmid pGW7-*pro*BA' showed similar protein profiles. Analysis of cells containing the plasmid pGW7-*pro*$(BA)_2$ revealed the plasmid-specified accumulation of two major protein species of molecular weight 37k and 41.5k. These protein bands were prominent even when protein was examined from cells grown at 30°C, and the intensity of staining increased further as the temperature of the culture was increased to 42°C. The appearance and accumulation of the two proteins were concomitant with the observed increases in enzymatic activity (Table 1), as well as with the appearance (as detected by SDS-polyacrylamide gel electrophoresis) of two ^{35}S-methionine labeled proteins of molecular weights of 37k and 41.5k. These proteins were synthesized preferentially during temperature induction of ^{35}S-methionine pulse-labeled cells harboring pGW7-*pro*$(AB)_2$ (data not shown).

A comparison of the electrophoretic mobilities of the purified GK and GPR enzymes with the total soluble protein extracted from temperature induced cells harboring pGW7-*pro*$(BA)_2$ was made by SDS-polyacrylamide gel electrophoresis. The induced protein of molecular weight 41.5k comigrated with purified GPR, and the induced protein of molecular weight of 37.0k comigrated with purified GK. The molecular weight of GPR is somewhat lower than that previously reported (18). Densitometric analysis of the Coomassie stained SDS-polyacrylamide gels indicated that GK and GPR together represented approximately 20-25% of the total soluble protein in the cell, with GPR accumulating in a 2 to 3-fold molar excess over GK (See Discussion).

Nucleotide Sequence of the 3.0 kb DNA Fragment Encoding GK and GPR

The strategy employed in obtaining the DNA sequence of the *E. coli proBA* region is shown in Figure 3. Analysis of the DNA sequence (Figure 4) revealed two tandem open reading frames encoding amino acid sequences that correlate well with those determined for the *proB* and *proA* gene products (Table 2, Figure 4). Each of the two open reading frames show a pattern of codon usage (data not shown) similar to that observed for the combined codon distribution of 52 *E. coli* genes (19).

The DNA sequence upstream from the 5'-terminus of the *proB* gene, from nucleotide position 150-214, is extremely AT-rich (86%) as compared to the *E. coli* genome (49% AT-rich). This region and flanking sequences contain multiple dyad symmetries (Figure 4).

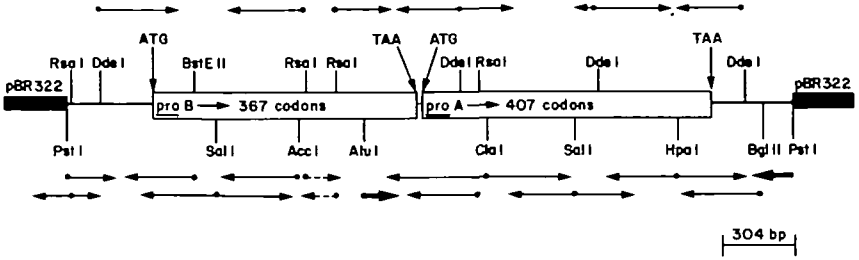

Figure 3. Sequencing Strategy for the proBA Region. The arrows denote sequencing direction and length of sequences obtained in individual experiments. Dashed arrows indicate sequence data obtained from strand-separated DNA fragments; bold arrows indicate sequence data obtained by the dideoxy chain-termination method; thin arrows indicate sequence data obtained by the Maxam and Gilbert method. Also shown are the restriction enzymes used to generate the various sequence start sites (•), as well as the codons for the initiation and termination of translation of the proB and proA gene products.

The nature of the DNA sequence upstream from the proB coding region precludes firm assignments of promoter regions. A 50 bp region encompassing the 5'-terminus of the proB gene was compared with published (20) E. coli promoter DNA sequences (Figure 5). The putative -10 region of the proB gene is homologous with the -10 region of the str and uvrB (P1) genes, and differs from the -10 region of several other genes by only a single nucleotide. The sequence ATTTTTC, which precedes the putative -10 region of proB, is homologous with the sequence present at the analogous region in the $T5-_{25}$ gene. The significance of this homology is uncertain since this region has not been implicated in any regulatory control of transcription. Finally, both λPRE and proB contain the sequence CGTT(Pyr)GTTTG, centered at position -35.

A putative Shine-Dalgarno sequence, GAG, precedes the proB initiator methionine codon and an open reading frame is terminated by a single TAA nonsense codon at position 1455 (Figure 4).

Two putative proA Shine-Dalgarno sites have been identified within the 14 bp intergenic region between proB and proA (Figure 4). The sequence TAAGGAG is complementary to the 3'-region of the 16S rRNA considered to be involved in the initiation of translation. This intergenic sequence presumably functions in the termination of translation of the proB gene product in addition to the initiation of translation of the proA gene product (see Discussion). An additional ribosomal binding site within the intergenic region might occur at AGG.

The proA gene encodes 407 amino acids and terminates at nucleotide

```
CTGCAGCCTGTACAGATGCAGATGCCACAATGCCCATCACCACTAATGCCAGAGTGCTCTTTTTCATTT      69
                                                                  ←phoE
TCATTCCTGATTTTAATTAACGCGCGAATATTCAGCGGGAGAGTCCCGTTGAAAACAGGAAAGTTTTTAAC   140
CTGAGATTGTTAAAGATATATTACAGATTAATAATATTCTTAAAATGTGGTAATTTATTAAATCTGTAATA   211
AAAGCGTAAACAACTGCCGCTAGGCTGCTGATCCCGCGCAACAAAACGCCATGCTTTGCTCGCAGATGGTT   282
GGCAACCGACGACAGTCCTGCTAAAACGTTCGTTTGATATCATTTTTCCTAAAATTGAATGGCAGAGAATC   353
                                                    -10
proB→
MET SER ASP SER GLN THR LEU VAL VAL LYS LEU GLY THR SER VAL LEU THR GLY
ATG AGT GAC AGC CAG ACG CTG GTG GTA AAA CTC GGC ACC AGT GTG CTA ACA GGC   407
                         Chi site

GLY SER ARG ARG LEU ASN ARG ALA HIS ILE VAL GLU LEU VAL ARG GLN CYS ALA
GGA TCG CGC CGT CTG AAC CGT GCC CAT ATC GTT GAA CTT GTT CGC CAG TGC GCG   461

GLN LEU HIS ALA ALA GLY HIS ARG ILE VAL ILE VAL THR SER GLY ALA ILE ALA
CAG TTA CAT GCC GCC GGG CAT CGG ATT GTT ATT GTG ACG TCG GGC GCG ATC GCC   515

ALA GLY ARG GLU HIS LEU GLY TYR PRO GLU LEU PRO ALA THR ILE ALA SER LYS
GCC GGA CGT GAG CAC CTG GGT TAC CCG GAA CTG CCA GCG ACC ATC GCC TCG AAA   569

GLN LEU LEU ALA ALA VAL GLY GLN SER ARG LEU ILE GLN LEU TRP GLU GLN LEU
CAA CTG CTG GCG GCG GTA GGG CAG AGT CGA CTG ATT CAA CTG TGG GAA CAG CTG   623

PHE SER ILE TYR GLY ILE HIS VAL GLY GLN MET LEU LEU THR ARG ALA ASP MET
TTT TCG ATT TAT GGC ATT CAC GTC GGG CAA ATG CTG CTG ACC CGT GCT GAT ATG   677

GLU ASP ARG GLU ARG PHE LEU ASN ALA ARG ASP THR LEU ARG ALA LEU LEU ASP
GAA GAC CGT GAA CGC TTC CTG AAC GCC CGC GAC ACC CTG CGA GCG TTG CTC GAT   731

ASN ASN ILE VAL PRO VAL ILE ASN GLU ASN ASP ALA VAL ALA THR ALA ALA ILE
AAC AAT ATC GTT CCG GTA ATC AAT GAG AAC GAT GCT GTC GCT ACG GCA GCG ATT   785

LYS VAL GLY ASP ASN ASP ASN LEU SER ALA LEU ALA ALA ILE LEU ALA GLY ALA
AAG GTC GGC GAT AAC GAT AAC CTT TCT GCG CTG GCG GCG ATT CTT GCG GGT GCC   839

ASP LYS LEU LEU LEU LEU THR ASP GLN LYS GLY LEU TYR THR ALA ASP PRO ARG
GAT AAA CTG TTG CTG CTG ACC GAT CAA AAA GGT TTG TAT ACC GCT GAC CCG CGC   893

SER ASN PRO GLN ALA GLU LEU ILE LYS ASP VAL TYR GLY ILE ASP ASP ALA LEU
AGC AAT CCG CAG GCA GAA CTG ATT AAA GAT GTT TAC GGC ATT GAT GAC GCA CTG   947

ARG ALA ILE ALA GLY ASP SER VAL SER GLY LEU GLY THR GLY GLY MET SER THR
CGC GCG ATT GCC GGT GAC AGC GTT TCA GGC CTC GGA ACT GGC GGC ATG AGT ACC  1001
```

```
LYS LEU GLN ALA ALA ASP VAL ALA CYS ARG ALA GLY ILE ASP THR ILE ILE ALA
AAA TTG CAG GCC GCT GAC GTG GCT TGC CGT GCG GGT ATC GAC ACC ATT ATT GCC   1055

ALA GLY SER LYS PRO GLY VAL ILE GLY ASP VAL MET GLU GLY ILE SER VAL GLY
GCG GGC AGC AAG CCG GGC GTT ATT GGT GAT GTG ATG GAA GGC ATT TCC GTC GGT   1109

THR LEU PHE HIS ALA GLN ALA THR PRO LEU GLU ASN ARG LYS ARG TRP ILE PHE
ACG CTG TTC CAT GCC CAG GCG ACT CCG CTT GAA AAC CGT AAA CGC TGG ATT TTC   1163

GLY ALA PRO PRO ALA GLY GLU ILE THR VAL ASP GLU GLY ALA THR ALA ALA ILE
GGT GCG CCG CCG GCG GGT GAA ATC ACG GTA GAT GAA GGG GCA ACT GCC GCC ATT   1217

LEU GLU ARG GLY SER SER LEU LEU PRO LYS GLY ILE LYS SER VAL THR GLY ASN
CTG GAA CGC GGC AGC TCC CTG TTG CCG AAA GGC ATT AAA AGC GTG ACT GGC AAT   1271

PHE SER ARG GLY GLU VAL ILE ARG ILE CYS ASN LEU GLU GLY ARG ASP ILE ALA
TTC TCG CGT GGT GAA GTC ATC CGC ATT TGC AAC CTC GAA GGC CGC GAT ATC GCC   1325

HIS GLY VAL SER ARG TYR ASN SER ASP ALA LEU ARG ARG ILE ALA GLY HIS HIS
CAC GGC GTC AGT CGT TAC AAC AGC GAT GCA TTA CGC CGT ATT GCC GGA CAC CAC   1379

SER GLN GLU ILE ASP ALA ILE LEU GLY TYR GLU TYR GLY PRO VAL ALA VAL HIS
TCG CAA GAA ATT GAT GCA ATA CTG GGA TAT GAA TAC GGC CCG GTT GCC GTT CAC   1433
                                              proA→
ARG ASP ASP MET ILE THR ARG ***               MET LEU GLU GLN MET GLY ILE
CGT GAT GAC ATG ATT ACC CGT TAA GGAGCAGGCTG ATG CTG GAA CAA ATG GGC ATT   1489

ALA ALA LYS GLN ALA SER TYR LYS LEU ALA GLN LEU SER SER ARG GLU LYS ASN
GCC GCG AAG CAA GCC TCG TAT AAA TTA GCG CAA CTC TCC AGC CGC GAA AAA AAT   1543
                        λatt
ARG VAL LEU GLU LYS ILE ALA ASP GLU LEU GLU ALA GLN SER GLU ILE ILE LEU
CGC GTG CTG GAA AAA ATC GCC GAT GAA CTG GAA GCA CAA AGC GAA ATC ATC CTC   1597

ASN ALA ASN ALA GLN ASP VAL ALA ASP ALA ARG ALA ASN GLY LEU SER GLU ALA
AAC GCT AAC GCC CAG GAT GTT GCT GAC GCG CGA GCC AAT GGC CTT AGC GAA GCG   1651

MET LEU ASP ARG LEU ALA LEU THR PRO ALA ARG LEU LYS GLY ILE ALA ASP ASP
ATG CTT GAC CGT CTG GCA CTG ACG CCC GCA CGG CTG AAA GGC ATT GCC GAC GAT   1705

VAL ARG GLN VAL CYS ASN LEU ALA ASP PRO VAL GLY GLN VAL ILE ASP GLY GLY
GTA CGT CAG GTG TGC AAC CTC GCC GAT CCG GTG GGG CAG GTA ATC GAT GGC GGC   1759
```

```
VAL LEU ASP SER GLY LEU ARG LEU GLU ARG ARG ARG VAL PRO LEU GLY VAL ILE
GTA CTG GAC AGC GGC CTG CGT CTT GAG CGT CGT CGC GTA CCG CTG GGG GTT ATT   1813

GLY VAL ILE TYR GLU ALA ARG PRO ASN VAL THR VAL ASP VAL ALA SER LEU CYS
GGC GTG ATT TAT GAA GCG CGC CCG AAC GTG ACG GTT GAT GTC GCT TCG CTG TGC   1867

LEU LYS THR GLY ASN ALA VAL ILE LEU ARG GLY GLY LYS GLU THR CYS ARG THR
CTG AAA ACC GGT AAT GCG GTG ATC CTG CGC GGT GGC AAA GAA ACG TGT CGC ACT   1921

ASN ALA ALA THR VAL ALA VAL ILE GLN ASP ALA LEU LYS SER CYS GLY LEU PRO
AAC GCT GCA ACG GTG GCG GTG ATT CAG GAC GCC CTG AAA TCC TGC GGC TTA CCG   1975

ALA GLY ALA VAL GLN ALA ILE ASP ASN PRO ASP ARG ALA LEU VAL SER GLU MET
GCG GGT GCC GTG CAG GCG ATT GAT AAT CCT GAC CGT GCG CTG GTC AGT GAA ATG   2029

LEU ARG MET ASP LYS TYR ILE ASP MET LEU ILE PRO ARG GLY GLY ALA GLY LEU
CTG CGT ATG GAT AAA TAC ATC GAC ATG CTG ATC CCG CGT GGT GGC GCT GGT TTG   2083

HIS LYS LEU CYS ARG GLU GLN SER THR ILE PRO VAL ILE THR GLY GLY ILE GLY
CAT AAA CTG TGC CGT GAA CAG TCG ACA ATC CCG GTG ATC ACA GGT GGT ATA GGC   2137

VAL CYS HIS ILE TYR VAL ASP GLU SER VAL GLU ILE ALA GLU ALA LEU LYS VAL
GTA TGC CAT ATT TAC GTT GAT GAA AGT GTA GAG ATC GCT GAA GCA TTA AAA GTG   2191

ILE VAL ASN ALA LYS THR GLN ARG PRO SER THR CYS ASN THR VAL GLU THR LEU
ATC GTC AAC GCG AAA ACT CAG CGT CCG AGC ACA TGT AAT ACG GTT GAA ACG TTG   2245

LEU VAL ASN LYS ASN ILE ALA ASP SER PHE LEU PRO ALA LEU SER LYS GLN MET
CTG GTG AAT AAA AAC ATC GCC GAT AGC TTC CTG CCC GCA TTA AGC AAA CAA ATG   2299

ALA GLU SER GLY VAL THR LEU HIS ALA ASP ALA ALA ALA LEU ALA GLN LEU GLN
GCG GAA AGC GGC GTG ACA TTA CAC GCA GAT GCA GCT GCA CTG GCG CAG TTG CAG   2353

ALA GLY PRO ALA LYS VAL VAL ALA VAL LYS ALA GLU GLU TYR ASP ASP GLU PHE
GCA GGC CCT GCG AAG GTG GTT GCT GTT AAA GCC GAA GAG TAT GAC GAT GAG TTT   2407

LEU SER LEU ASP LEU ASN VAL LYS ILE VAL SER ASP LEU ASP ASP ALA ILE ALA
CTG TCA TTA GAT TTG AAC GTC AAA ATC GTC AGC GAT CTT GAC GAT GCC ATC GCC   2461

HIS ILE ARG GLU HIS GLY THR GLN HIS SER ASP ALA ILE LEU THR ARG ASP MET
CAT ATT CGT GAA CAC GGC ACA CAA CAC TCC GAT GCG ATC CTG ACC CGC GAT ATG   2515

ARG ASN ALA GLN ARG PHE VAL ASN ALA SER THR ARG PHE THR ASP GLY GLY GLN
CGC AAC GCC CAG CGT TTT GTT AAC GCC TCT ACG CGT TTT ACC GAC GGC GGC CAG   2569
```

```
PHE GLY LEU GLY ALA GLU VAL ALA VAL SER THR GLN LYS LEU HIS ALA ARG GLY
TTT GGT CTG GGT GCG GAA GTG GCG GTA AGC ACA CAA AAA CTC CAC GCG CGT GGC   2623

PRO MET GLY LEU GLU ALA LEU THR THR TYR LYS TRP ILE GLY ILE GLY ASP TYR
CCA ATG GGG CTG GAA GCA CTG ACC ACT TAC AAG TGG ATC GGC ATT GGT GAT TAC   2677

THR ILE ARG ALA ***                                    ***▼
ACC ATT CGT GCG TAA ATAAAACCGGGTGATGCAAAAGTAGCCATTTGATTCACAAGGCCATTGACG   2743

CATCGCCCGGTTAGTTTTAACCTTGTCCACCGTGATTCACGTTCGTGAACATGTCCTTTCAGGGCCGATAT   2814

AGCTCAGTTGGTAGAGCAGCGCATTCGTAATGCGAAGGTCGTAGGTTCGACTCCTATTATCGGCACCATTA   2885

AAATCAAATTGTTACGTAAGATCTTATCATTCTCCCACCAAAAAATTATCTTAATGTAACAGCTGGTGTAA   2956

GTAAATTCTATCAACGAAGATCAATCTTATCTACTGACCAAAAAGGCCTGATAGGGCTTCGCTCACTATAC   3027

ATCCTTGGCTGCAG                                                             3041
```

Figure 4. Nucleotide Sequence of the 3.04 kb PstI DNA Fragment Containing the proBA Region. The 5'-end of the phoE gene (sense strand) is shown including the position of a putative Shine and Dalgarno sequence (underlined). Regions of dyad symmetry [axes indicated by (▼)] thought to be involved in the transcriptional control of either the phoE gene or the proB gene are indicated. The putative -10 region (dashed line) of the proB gene is indicated along with the putative Shine and Dalgarno sequences (underlined) and translation termination codons (asterisks) for the proB and proA genes. A region of dyad symmetry thought to be involved in the termination of transcription of the proA gene is indicated as well as the proposed stop-site of transcription (dashed line). Confirmation of a portion of the amino acid sequence deduced from the DNA sequence was obtained by NH_2- terminal and COOH-terminal sequence analyses of purified γ-glutamyl kinase (proB) and γ-glutamyl phosphate reductase (proA) protein and is indicated by overlines.

position 2690 with TAA. A second nonsense codon, TGA, occurs further downstream, in phase with the first nonsense codon. This second codon is located within an inverted repeat sequence centered at position 2726. A possible RNA transcription termination structure with features characteristic of rho-independent terminators (19) is shown in Figure 6.

The secondary attachment site for bacteriophage lambda had already been sequenced (22), and mapped by complementation (23) to the proBA region. The sequence data presented here shows that the secondary attachment site (GTATAAA) resides in the N-terminal portion of the proA gene (Figure 4).

The calculated molecular weight for the GK (proB) and GPR (proA) monomers based upon the DNA sequence was 38,952 and 43,503, respectively. These values are consistent with those obtained from SDS-polyacrylamide gel electrophoretic analysis of total cellular protein (Figure 2). Based upon a molecular weight (9) of 236,000 and 268,000 for the holoenzymes of GK and GPR, respectively, GK

TABLE 2. Amino acid composition of the GK and the GPR proteins

Amino Acid	GK Analysis Protein[a]	GK Analysis DNA[b]	GPR Analysis Protein[a]	GPR Analysis DNA[b]
non-polar				
Ala	46	46	58	53
Gly	35	37	34	31
Ile	29	31	22	27
Leu	45	39	48	43
Met	4	6	8	9
Phe	7	5	2	5
Pro	10	11	20	12
Val	30	25	35	35
		200		215
polar				
Asn	n.d.	13	n.d.	16
Cys	1	3	n.d.	7
Gln	n.d.	13	n.d.	18
Ser	21	22	16	20
Thr	18	19	16	21
Trp	n.d.	2	n.d.	1
Tyr	8	7	5	7
		81		90
acidic				
Asp (Asx)	(40)	24 (37)	(46)	28 (44)
Glu (Glx)	(35)	17 (30)	(46)	23 (41)
		39		51
basic				
Arg	26	26	21	25
His	8	10	6	7
Lys	8	11	15	19
		47		51

[a]Data are amino acid residues, as determined by amino acid analysis, to the nearest whole number.
[b]Composition determined from DNA sequence.
n.d., not determined.

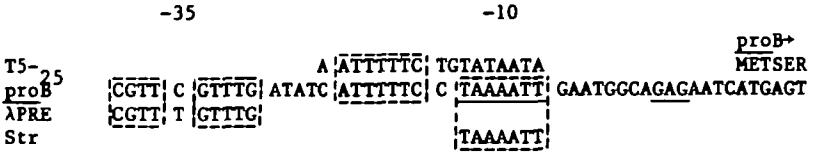

Figure 5. Homologies in the Nucleotide Sequences of the 5'-Regulatory Region of proB. The 5' sequences of the proB gene are compared with the -10 sequences of Str and λPRE and the -35 sequence of T5-25. Underlined and boxed sequences are explained in text.

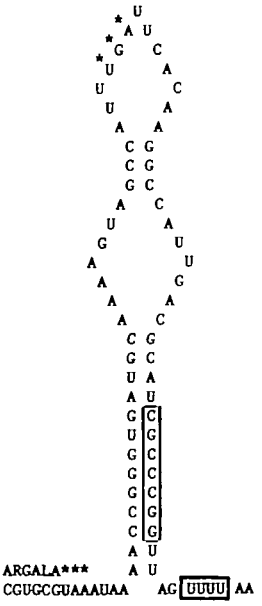

Figure 6. A Possible RNA Transcription Termination Structure at the 3'-End of the proA Gene. The 3'-end of the proA gene has been represented as a sequence of ribonucleotides in a stem and loop secondary structure. The COOH-terminal alanine codon (GCG) is followed by both a UAA nonsense codon (***) and, further downstream and in phase, a UGA nonsense codon (***). A GC-rich sequence present in the stem, followed by a U-rich region are boxed. The termination structure has been drawn to maximize base pairing stability according to the rules of Tinoco et. al. (35). The stem has a ΔG value of -33.4 kcal while the entire stem and loop structure has a ΔG value of -31.4 kcal.

and GPR are each comprised of 6 identical subunits.

Amino Acid Sequence of GK and GPR

The complete amino acid sequence of GK and GPR is shown in Figure 4. NH_2 and COOH-terminal amino acid analysis of both purified enzymes corresponded with that predicted by the DNA sequence (Figure 4). Further confirmation of the putative protein sequences was provided by the amino acid composition obtained for both purified proteins which correlated well with that predicted from the DNA sequence (Table 2).

DISCUSSION

In this study the genes encoding GK (proB) and GPR (proA), two of the three enzymes in the proline biosynthetic pathway, were subcloned, identified by protein and DNA sequence analysis, and placed under the control of the highly active λP_L promoter. Though the proB and proA loci had been placed at about 6 min. on the E. coli linkage map (24), it was unclear whether we would find the proB and proA genes arranged contiguously (25) or separated by intervening genes (26, 27, 28). Complementation and enzymology studies employing plasmids obtained from the Clarke and Carbon collection of hybrid plasmids indicated that the proB and proA genes were within at least 20 kb of each other (1). Recently, several laboratories (10,29,30,31) have identified

a 3 kb E. coli DNA fragment which encodes GK and GPR activity. With this present report we now document that the proB and proA genes are indeed contiguous and very likely comprise a polycistronic operon. However, we cannot rule out the possibility (as some of the above studies might suggest) of genomic fluidity in the proBA region. During our analysis of the DNA sequence data obtained for the proBA region, we discovered the presence of a sequence within proB, GCTGGTGG (Figure 4), homologous to a chi site - a recombinational hotspot around which the rate of RecBC-promoted recombination is elevated (21). If the proBA region is subject to genomic fluidity then the presence of a sequence homologous to a hotspot of generalized recombination might explain the occurrence of strain specific gene rearrangements in the proBA region.

In the present study we found no evidence of a DNA sequence upstream from the 5'-end of the proB gene that corresponded to a transcription attenuator. Previously published results (6) had indicated that the synthesis of GK was repressed by proline. However, we as well as others (10,1) have found that the specific activity of GK is not affected by the addition of proline to the growth medium. In addition, the regulation of enzyme synthesis in proBA-lacZ fusion strains was found not to be influenced by the type of carbon source (i.e., proline or glucose) in the growth medium (10).

Several inverted repeats were identified upstream from the 5'-end of the proB gene (see Figure 4). During the course of this research we had identified, via a promoter assay vector, a promoter in close proximity to the proB promoter that stimulated transcription from the opposite strand (Don VanLeeuwen, unpublished results). A recent report (32), detailing the nucleotide sequence of the E. coli phoE gene, positioned the phoE gene on the opposite strand and adjacent to the proA gene. While our DNA sequence of this region agrees with that reported, it is clear from our research that the phoE gene actually abutts the proB gene (Figure 4). The close proximity of the phoE and proB 5'- transcription regulatory regions makes structural/functional assignments for the inverted repeats difficult. Apparently both the phoE and proB genes lack the canonical -35 sequence.

The concept of the proBA locus functioning as an operon (30,31) is supported by our observations that 1) analysis of the DNA sequence preceding the proA coding region failed to reveal attractive promoter sequences, 2) only 14 bp separate the COOH-terminal of proB from the NH_2-terminal of proA, with the proB termination codon most likely overlapping with the proA ribosomal binding site, and 3) the absence at the 3'-end of the proB gene of visible homology to a rho-independent terminator.

A rationale for the polycistronic nature of the proBA locus can also be constructed based upon an enzymological argument. The product of the reaction catalyzed by GK is a labile intermediate, γ-glutamyl phosphate, and it has been suggested (3) that GK and GPR form a molecular complex which ensures the direct transfer of γ-glutamyl phosphate from GK to GPR. The biosynthesis of both enzymes in close spacial proximity to one another would result in relatively high localized concentrations of both enzymes and this would facilitate the formation of a physical complex. In support of the existence of this presumptive complex, we have found that purified preparations of GK are enzymatically inactive in enzyme assays employing hydroxymate, but are enzymatically active following the addition of a homogenous preparation of GPR (9).

The existence of this enzymatic complex would lead to the requirement that cellular levels of GK and GPR either be tightly regulated in a stoichiometric fashion, or with GPR being in molar excess over GK. Several lines of evidence indicate that GPR is synthesized in molar excess: (1) Hayzer reported (1) an almost 6-fold differential increase in the specific activity of GPR in comparison to GK in cells transformed with a ColEI hybrid plasmid containing the proBA region. (2) In this present communication we report a 2-fold differential increase in specific activity in favor of GPR when the proBA genes were cloned into the expression plasmid pGW7, and transformed cells were grown at 30°C or 37°C (Table I). At 42°C, with the λP_L promoter fully induced, we observed a 3-fold differential increase in specific activity in favor of GPR. (3) SDS-polyacrylamide gel electrophoresis also revealed a corresponding differential increase in the amounts of the two enzymes (Figure 2). Thus, it would appear that the synthesis of GPR is regulated in a fashion that ensures a molar excess of GPR over GK.

Transcriptional and translational regulatory mechanisms that could conceivably function to ensure either stoichiometric or molar excess amounts of GPR become apparent upon examination of the proBA DNA sequence. First, the polycistronic nature of the proBA genes would facilitate the synthesis of stoichiometric amounts of mRNA sequences encoding both proteins. Second, the overlap of the proB translation termination codon with the proA ribosomal binding site could allow translational coupling (33) i.e., the maintenance of ribosomal association with the mRNA throughout the translation of both protein encoding mRNA sequences. Translational reinitiation without ribosome release would help to ensure stoichiometric production of both polypeptides. Third, the potential availability of several ribosomal binding sites (one in proB and

two in proA) directing ribosomes to the GPR mRNA sequence would also ensure a molar excess of GPR. In addition, the extensive homology between the proA ribosomal binding site, TAAGGAG (versus GAG for proB), and the 3'-end of the 16S rRNA potentially portends an advantage in the competition for ribosomes.

The three genes coding for the enzymes of the proline biosynthetic pathway have been completely sequenced and characterized (7, present report). With the knowledge obtained from our studies we have constructed several vectors enabling the overproduction of proline in industrially significant quantities by strains of E. coli (10). In addition, proline is a primary osmoregulatory agent (36) whose overproduction in agriculturally significant organisms via genetic engineering technology might increase their osmotic tolerance and therefore their environmental tolerance and productivity. The ability to manipulate the expression of the enzymes of the proline biosynthetic pathway should result in a better understanding of osmoregulation in bacteria and animals as well as in plants. Finally, comparison of our sequence data of the mutant proB gene, responsible for conferring the DHP analog resistant phenotype, with the DNA sequence of the wild-type proB gene (manuscript in preparation) should facilitate the elucidation, and optimization by site-specific mutagenesis, of the structural and molecular interactions determining proline feedback resistance.

ACKNOWLEDGEMENTS

We thank Lisa Bowers and Garth Ehrlich for their technical assistance during the initial stages of this investigation. We also thank Dr. Nancy Sasavage for cloning the 3.0 kb PstI DNA fragment into M13-mp8 and for providing us with additional sequence data. We are grateful to Dr. Jennifer Pinkham (Yale University) for sharing with us her unpublished and confirming sequence information flanking the chi site. We are indebted to Dr. Donald Capra for providing the amino acid analysis data. Expert and patient typing of the manuscript was performed by Patti Restivo and Lynne Keady.

[1]Present address: Washington Research Center, W.R.Grace & Co., 7379 Route 32, Columbia, MD 21044, USA

[2]Present address: Syngene Products & Research, 225 Commerce Drive, P.O. Box 2211, Fort Collins, CO 80522, USA

[3]Present address: Laboratory of Molecular Microbiology, NIAID, Building 550, Fort Detrick, Frederick, MD 21701, USA

*To whom correspondence should be addressed

REFERENCES

1. Hayzer, D. J. and Leisinger, Th. (1980) J. Gen. Microbiol. 118, 287-293.
2. Baich, A. and Pierson, D. J. (1965) Biochem. Biophys. Acta 104, 397-404.
3. Baich, A. (1969) Biochem. Biophys. Acta 192, 462-467.
4. Baich, A. (1971) Biochem. Biophys. Acta 244, 129-134.
5. Rossi, J. J., Verder, J., Berg, C. M. and Coleman, W. H. (1977) J. Bacteriol. 129, 108-114.
6. Tristram, H. and Thurston, C. F. (1966) Nature 212, 74-75.
7. Deutch, A. H., Smith, C. J., Rushlow, K. E. and Kretschmer, P. J. (1982) Nucleic Acid Res. 10, 7701-7714.
8. Clarke, L. and Carbon, J. (1975) Cell 9, 91-99.
9. Smith, C. J., Deutch, A. H. and Rushlow, K. J. (1984) J. Bacteriol. 157, 545-551.
10. Bloom, F., Smith, C. J., Jessie, J., Veilleux, B. and Deutch, A. H. (1984) in Fifteenth Miami Winter Symposium, Vol. 20 (in press).
11. Hoopes, B. C. and McClure, W. R. (1981) Nucleic Acid Res. 9, 5493-5504.
12. Williams, I. and Frank, L. (1975) Anal. Biochem. 64, 85-97.
13. Maxam, A. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K., Vol. 65, pp. 499-560, Academic Press, New York.
14. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A. (1980) J. Mol. Biol. 143, 161-178.
15. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
16. Klapper, D. G., Wilde, C. E. and Capra, J. D. (1978) Anal. Biochem. 85, 126.
17. Karp, D. R., Parker, K. L., Schreffler, D. C. and Capra, J. D. (1981) 126, 2060.
18. Hayzer, D. J. and Leisinger, Th. (1982) Eur. J. Biochem. 121, 561-565.
19. Alff-Steinberger, C. (1984) Nucleic Acid Res. 12, 2235-2241.
20. Hawley, D. K. and McClure, W. R. (1983) Nucleic Acid Res. 11, 2237-2255.
21. Smith, G. R., Kunes, S. M., Schultz, D. W., Taylor, A. and Triman, K. L. (1981) Cell 24, 429-436.
22. Pinkham, J. L., Platt, T., Enquist, L. W. and Weisberg, R. A. (1980) J. Mol. Biol. 144, 587-592.
23. Shimada, K., Weisberg, R. A. and Gottesman, M. E. (1972) J. Mol. Biol. 63, 483-503.
24. Bachmann, B. J. (1983) Microbiological Rev. 47, 180-230.
25. Condamine, H. (1971) Annales de l'Institut Pasteur 120, 9-22.
26. Broda, P. (1974) J. Bacteriol. 117, 741-746.
27. Roberts, L. M. and Reeve, E. C. R. (1970) Genetical Research 16, 359-365.
28. Havekes, L. M., Lugtenberg, B. J. J. and Hoekstra, W. P. M. (1976) Molec. Gen. Genet. 146, 43-50.
29. LeRudulier, D. and Valentine, R. C. (1982) Trends in Biochemical Sciences 7, 431-433.
30. Hayzer, D. J. (1983) J. Gen. Microbiol. 129, 3215-3225.
31. Mahan, M. J. and Csonka, L. N. (1983) J. Bacteriol. 156, 1249-1262.
32. Overbecke, N., Bergman, H., Mansfeld, F. van and Lugtenberg, B. (1983) J. Mol. Biol. 163, 513-532.
33. Oppenheim, D. S. and Yanofsky, C. (1980) Genetics 95, 785-795.
34. Laemmli, U. K. (1970) Nature 227, 680-685.
35. Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. and Gralla, J. (1973) Nature New Biology 246, 40-41.
36. LeRudulier, D., Strom, A.R., Dandekar, A.M., Smith, L. T. and Valentine, R. C. (1984) Science 224, 1064-1068.